Application of Clustering for the Development of Retrofit Strategies for Large Building Stocks

Philipp Geyer[1], Arno Schlüter[2], Sasha Cisar[2]
[1] KU Leuven, Belgium, p.geyer@kuleuven.be
[2] ETH Zürich, Switzerland

Abstract. In order to reduce energy consumption and emissions from the built environment, it is vital to transform the existing building stock and develop retrofit strategies to achieve energy efficiency and building-integrated renewable energy supply. Compared to developing cost-optimal retrofit strategies for one building, the development of strategies for 100 to up to 10,000 buildings remains a major challenge. This paper presents a method to cluster buildings based on their sensitivity to different retrofit measures, focusing on the cost-effectiveness. Derived from algorithmic clustering and combined with time and cost data, a tailored development of retrofit strategies for large building stocks becomes possible. Improved identification of retrofit measures and strategies, in contrast to the conventional classification based on building type and age, is demonstrated. The method is illustrated, using the data from the case study project 'Zernez Energia 2020', which aims to achieve carbon neutrality of a Swiss alpine village.

Keywords: Algorithmic Clustering, Building Retrofitting, Strategic Building Stock Management, Data Mining.

1. Introduction

Achieving the objective of energy efficiency and emission reduction in urban structures requires the management of retrofitting the building stock. A central need is the identification of the energy reduction potential for heating, cooling and related emissions and to locally produce renewable energy within the constraints of a limited financial budget. To achieve an optimal retrofit it would be beneficial to assess each single building in an urban structure for the applicability and effect of measures for efficiency improvement and decentralized energy production and then to draw an overall conclusion about the potential. Furthermore, as building owners need to finance the largest portion of the measures themselves, it is essential to identify strategies about how to best retrofit and improve the energetic behaviour of the buildings in relation to the available means of investment.

The currently most applied approach to deal with building stocks of 100 up to 10,000 buildings is a type-age classification. However, for the determination of measures for improving energy efficiency and for reducing CO_2 emissions, many other factors than just age and building type have an impact on the effectiveness of retrofit measures. Therefore, the type-age classification provides only a rough estimation to develop retrofit strategies. Using modern databases and GIS, comprehensive information of the building stock is increasingly available and this paves the way to develop more sophisticated approaches that utilize this information allowing retrofit strategies to be implemented in a timelier, efficient and cost effective manner.

1.1 Objectives

The objective of this work is to utilize available rich building data for simulation, analysis and identification of cost-optimal retrofits measures for groups of individual buildings. The resulting matching of measures to building groups can then be utilised to define overarching strategies in order to allow faster and more cost-effective retrofits. The development of strategies for clusters instead of individual buildings facilitates more effective strategy development, compared to dealing with each building individually, and makes it thus more feasible to address the complete building stock. Such information can lead to a change in policies for retrofit subsidies and a more effective utilization of public funds. Furthermore, a coordinated retrofit opens up possibilities of the economy of scale.

We present the method of performance-based clustering in Section 2. This includes a comparison of different clustering techniques. The case study of Zernez, which is introduced in Section 3, is used to demonstrate that the proposed method improves the impact of retrofit measures while minimizing the resource consumption (material and costs). The method is applied on the data collected during the case study and related available information. Section 4 reports on the result of the application of the method on the case study. This section also shows how to derive similar reacting groups of buildings, how to interpret them and how to develop retrofit strategies.

1.2 Background

For predicting the future energy use and emissions and for evaluating scenarios and strategies, building stock models are important. Swan and Ugursal (2009) and Kavgic et al. (2010) review the state of the art techniques of building stock modelling in research, existing models and approaches and their use in policy making. The typical parameters that these models build on are categories of building age, type of building, heat distribution type, energy source, construction or refurbishment year and dwelling type. In particular, bottom-up models that are based on building physics engineering serve to assess the reduction potential of energy efficiency measures and technologies for the building stock. For this purpose, existing approaches usually define scenarios and strategies using standard combinations of measures and estimation for the whole building stock. However, these approaches do not derive measures or strategies from the analysis of the actual building stock and do not derive individual retrofit strategies.

Furthermore, classification methods are applied to building stock data. Usually, the parameters in the data serve to derive groups and to assign measures for energy efficiency. A typical example is the type-age classification that has been developed for Germany (Kohler et al. 1999) and recently extended to Europe (IWU, 2012). Many examples of classifying are based on type, age and other similar parameters in building stock. For instance, Kohler and Yang (2007) address long term management of building stocks by using the German model of Building Stocks. Uihlein and Eder (2010) use four age categories in the development of a broad strategy for EU-27 residential building stock. Boardman (2007) examines low- and zero-carbon technologies for residential building stock in UK. He uses Oxford’s UK Domestic Carbon

Model that is based on age class, dwelling type, tenure type, number of floors and construction type.

Moreover, approaches that use algorithmic clustering for building stocks based on energy consumption and other parameters exist. Santamouris et al. (2007) apply clustering to a database of 320 schools in Greece and build groups based on the energy consumption with climatic normalization. Gaitani et al. (2010) identify typical building properties and parameters of the schools by k-means clustering. Jones et al. (2001) cluster building stock by building properties, such as heated ground floor area, facade, window to wall ratio. Yamaguchi et al. (2007) identify district types and provide typical energy performance by simulating buildings in a representative district.

Furthermore, in engineering fields other than building stock management, de Oliveira et al. (2011) use divisive hierarchical clustering of failures in a water network. Jazizadeh et al. (2014) apply heuristic unsupervised hierarchical clustering for autonomous partitioning of appliances signature space in non-intrusive load monitoring (NILM). Motamedi et al. (2013) use spatial clustering combined with other criteria during localization of RFID-equipped assets. Hung and Kang (2014) develop a method for grouping objects for collision detection in real-time construction simulation using k-means clustering. Hyun et al. (2015) use hierarchical clustering for similarity analysis of car designs in order to identify car manufacturers' design strategies.

All mentioned approaches related specifically to building stocks use either the descriptive parameters (age, type, etc.) or the current performance to set up groups of buildings. Only a very few, such as Gaitani et al. (2010) explicitly use algorithmic clustering. However none of the existing methods apply algorithmic clustering based on effectiveness of measures to develop retrofitting strategies for a building stock. This is of major interest for the development of strategies, as it is much easier to make decisions and develop strategies for groups of buildings that react similarly to energy efficiency measures.

2. The method of performance-based clustering

In this section, we describe the methodology for the fundamental shift from clustering based on describing parameters to clustering based on performance-based indicators of measure effects. This change allows identification of groups that react similarly to energy efficiency measures and therefore form the basis for group strategies. The main innovation of the method consists of selecting appropriate performance indicators for clustering that individually describe the reaction of retrofit measures instead of parameters such as type and age or general energy consumption. This is the basis for deriving groups that react similarly on a planning strategy developed for the group. Figure 1 provides an overview of the method. The four major phases of the process are: (1) data preparation, (2) pre-processing, (3) clustering and (4) post-processing.

2.1 Data preparation

The aim of the data preparation is to develop a database that includes all buildings with the describing parameters that are required to estimate the effect of energy efficiency measures.

Important parameters include the total and heated floor area and the use of the buildings as well as external surface of façade, windows and doors, roof, walls etc. Orientation, geometry and available roof area play an important role for solar collectors (photovoltaic, PV and solar thermal, ST). Moreover, the existing heating system and the consumed fuel or electric power per year are valuable information. Finally, the setup of a GIS model was a helpful means in the Zernez project for interpreting results, for spatial analysis and for proposing district networks.

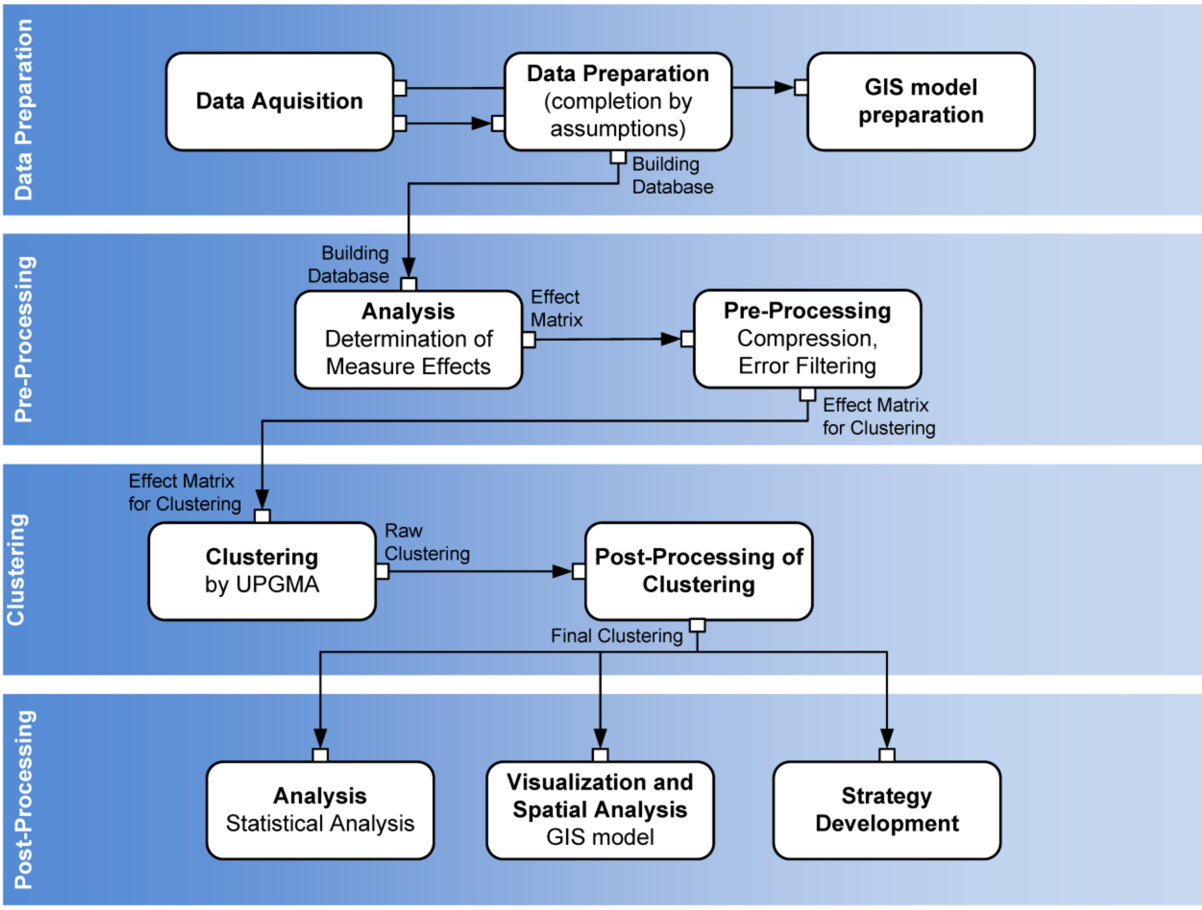

Figure 1. The flow chart shows the method of determining clusters and strategies from the given building stock database.

2.2 Pre-processing

Defining performance-based feature space for clustering

The first step is the definition of the dimensions of the feature space for clustering. This definition leads to j features characterizing a building instance. These features consist either of the effect of a simple energy efficiency (EE) measure, for instance, insulation of the building envelope or the effect of a measure set, for instance, insulation of the envelope and new heating system. For each feature, the effect e of the measures or the measure set is determined as the quotient of emission reduction ΔV_{CO2} and the investment costs c_{invest}:

$$e_j = \frac{\Delta V_{CO_2,\, j}}{c_{invest,\, j}} \quad \text{in g } CO_2\text{/a per CHF}. \qquad (1)$$

Alternatively, clustering, using only CO_2 reductions, was performed. This delivers information insensitive to costs for cases where only the reduction of emission are considered without analysis of any investment costs.

Determining the effect matrix for clustering

The basis for clustering is the response of each building to each measure, which defines the matrix *E*. This matrix locates every building in the virtual feature space defined in the previous step. For the building *i* and the measure *j*, the effect matrix composes as follows:

$$\mathbf{E} = e_{i,\ j}\,. \tag{2}$$

To determine the effect of a measure, a calculation method is required determining the reduction of CO_2 emissions and the investment costs per building. Fundamentally, a significant number of different methods are available. For energy calculations, these could be simplified estimations, monthly sums or dynamic simulation. The main requirement is that the calculation methods scale well for the large number of cases and features. Section 4.1 presents the simplified estimation applied in the case study. It is important to note that the clustering only relies on comparative information of the performance calculation not on the absolute values.

Compression of the effect matrix

Before applying a distance-based clustering to this space, compression is applied. This avoids clusters with only a few runaways. Such runaways occur in case of buildings with special situations, such as a very low energy consumption and high potential to produce energy. Compression avoids clusters with only a few runaways that often occur because of special buildings in a stock. For the compression a logarithmic transformation of the unit-less effect matrix *E* serves:

$$\mathbf{E}_{comp} = \log\left(e_{i,j} + 1\right), \tag{3}$$

which generates the compressed effect matrix E_{comp}.

2.3 Clustering

Having the feature space defined by the previous steps, algorithmic clustering is applied to the instances in the feature space. A clustering method serves to identify groups of instances closely located to each other in the feature space. The next subsection characterizes and examines available methods for this purpose.

After clustering, interpretation and manual adjustments are required. The expert interprets results and, where necessary, performs minor redistributions of instances. The main tool for this purpose is scatter plotting that pairwise shows the location of instances in the feature space. As usually feature spaces of retrofit decisions have multiple measures, one has to handle a number of scatter plots at the same time, such as shown in Figure 7.

Introduction to clustering

From a mathematical point of view: “Clustering is the process of grouping together objects that are similar” (Simovici and Djeraba 2008, 495). In the presented study, this characteristic of an algorithmic clustering method is applied to identify instances reacting similarly to retrofit measures.

The basic function of clustering is to assign those instances to one cluster or group that are close in the feature space. Therefore, the measure of the distance is the basis for clustering. Furthermore, a procedure for assigning instances to clusters is required. Due to its simplicity and deterministic character, agglomerative hierarchical clustering was selected. For comparison purposes only, partitioning k-means clustering was also considered.

(1) **Hierarchical clustering:** Agglomerative hierarchical clustering according to Xu and Wunsch (2009) uses the following steps:

- Assign each instance to one individual cluster. All clusters have one instance;
- Calculate the proximity matrix of the clusters;
- Merge the two closest clusters;
- Repeat the steps 1 to 3 until the desired number of clusters is reached.

Furthermore, the methods differ in the way the proximity matrix is calculated. The case study uses the implementation of the Statistical Toolbox of MATLAB. The following list shortly describes the methods considered as candidates for performance-based clustering:

a. **Unweighted Pair Group Method with Arithmetic mean (UPGMA)**: This method is attributed to Sokal and Michener (1958). It uses the average distance of all cluster members:

$$d(r,s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} dist\left(x_{r,i}, x_{s,j}\right). \tag{4}$$

b. **Unweighted Pair Group Method using Centroids (UPGMC)**: This method uses the Euclidian distance between the two clusters' centroids as linkage:

$$d(r,s) = \left\| \bar{x}_r - \bar{x}_s \right\|_2 \text{ with the average } \bar{x}_r = \frac{1}{n_r} \sum_{i=1}^{n_r} x_{r,i}. \tag{5}$$

c. **Weighted Pair Group Method using Centroid (WPGMC)**: This method stores the cluster centroids $\tilde{x}$ and calculates the difference between them:

$$d(r,s) = \left\| \tilde{x}_r - \tilde{x}_s \right\|_2. \tag{6}$$

In case of two clusters being merged, the new centroid is calculated as follows and stored for the cluster:

$$\tilde{x}_r = \frac{1}{2}\left(\tilde{x}_p + \tilde{x}_q\right). \tag{7}$$

For WPGMC and the next method (WPGMA) the ‘weighting’ aspect means that the points are not considered equally but by merging cluster centroids implicit weight is put to clusters with less points.

d. **Weighted Pair Group Method with Arithmetic mean (WPGMA)**: This method determines the distance to a third cluster when merging two clusters by the mean of both clusters' distance to the third one:

$$d(r,s) = \frac{d(p,s) + d(q,s)}{2}. \tag{8}$$

e. **Shortest Distance (SD):** This method uses the closest members from two clusters to determine the distance:

$$d(r,s) = \min\left(dist\left(x_{r,i}, x_{s,j}\right)\right) \text{ with } i \in (1,...,n_r),\ j \in (1,...,n_s). \tag{9}$$

(2) **Partitioning clustering:** In contrast to hierarchical clustering, partitioning clustering does not merge clusters with other clusters. Instead, the instances are assigned to clusters considering an objective criterion.

a. **K-Means-Method**: This method uses the minimization of the inner squared distance sum of a cluster as criterion.

$$\text{minimize } D = \sum_{i=1}^{n_c} \sum_{x \in K_i} \left(x - \overline{x}_i\right)^2 \tag{10}$$

Forgy (1965) and Lload (1982) developed the standard methods for this minimization.

Because of their simplicity and effectiveness, the methods described above represent the most extensively used ones among the multiple available clustering methods as described in Xu and Wunsch (2009).

Comparative testing of clustering methods

To learn more about the influence of the clustering method, hierarchical and partitioning algorithms are compared in a test. This comparison uses the data from the case study presented in the Sections 3 and 4. The CO_2 conversion factors is set for the European electricity network and the number of clusters is fixed to seven. The UPGMA method serves as reference. Criterion for comparison is the deviation in assigning the buildings to the clusters by the different clustering methods.

Due to the different procedures of the algorithms, the order of the identified clusters is not identical and a re-ordering of the clusters was required. For this purpose, an auxiliary algorithm was developed in order to reduce the number of deviating classification of buildings in clusters compared to the reference. The results to that this algorithm has been applied are marked by an ‘R’. The algorithm has the following structure in pseudo code:

```
DO for all test clusters
   Find the most frequent pair of test/reference cluster number
```

```
    Replace the number in a copy of the test clustering by the
        number of the reference in this pair
END
```

Figure 2 shows the results of the comparison based on the re-ordered clusters. Reference for the comparisons are seven clusters derived by the UPGMA method, whose number of members is shown by the large dark blue bars. The small coloured difference bars show the number of deviating classifications by other clustering methods that are not avoidable by re-ordering.

These results show only slight difference between the two different categories of clustering methods, i.e. hierarchical and partitioning clustering. For example, the green bar shows the difference between UPGMA as a hierarchical method to k-means clustering as a partitioning method. However, the aggregation method of the points for the fusion of clusters, i.e. weighted or unweighted averaging or shortest distance between individuals, caused major differences. The unweighted, which includes k-means, against the weighted method leads to the biggest difference. Weighting in this context means that the centroids are not calculated from the average of all instances in the cluster but they are passed on during the fusion of clusters; the average of the previous two centroids forms the new centroid, not the average of the single points. This means an implicit higher weight for clusters with less instances. Also the SD method has a high deviation from the non-averaging methods. Therefore, it is recommended that the focus is not on the type of clustering algorithm for identifying consistent and sound clusters but instead on the aggregation method in these algorithms.

The choice of unweighted hierarchical clustering for the later application in the case study relies on good conformity of the unweighted hierarchical methods and k-means as partitioning method. However, the method of performance-based clustering is compatible with different clustering methods.

2.4 Post-processing

Given the final clustering, before developing a strategy, different analyses and interpretations were undertaken to support the understanding of the results. A key instrument in this case are scatterplots illustrating the location of clusters in the feature space. Special attention is paid to differences in the reaction of clusters to measures. This is fundamental information for developing strategies per group. For instance, in the case study, statistics showed that clusters with highest priority for action as having high emissions include mainly oil heating systems and that these buildings were built mainly in the second half of the 20th century. Furthermore, spatial analysis using GIS technology shows the location of clusters for the development of district energy system, such as district heating or thermal microgrids—scenarios for which spatial distribution of energy demand and supply are crucial.

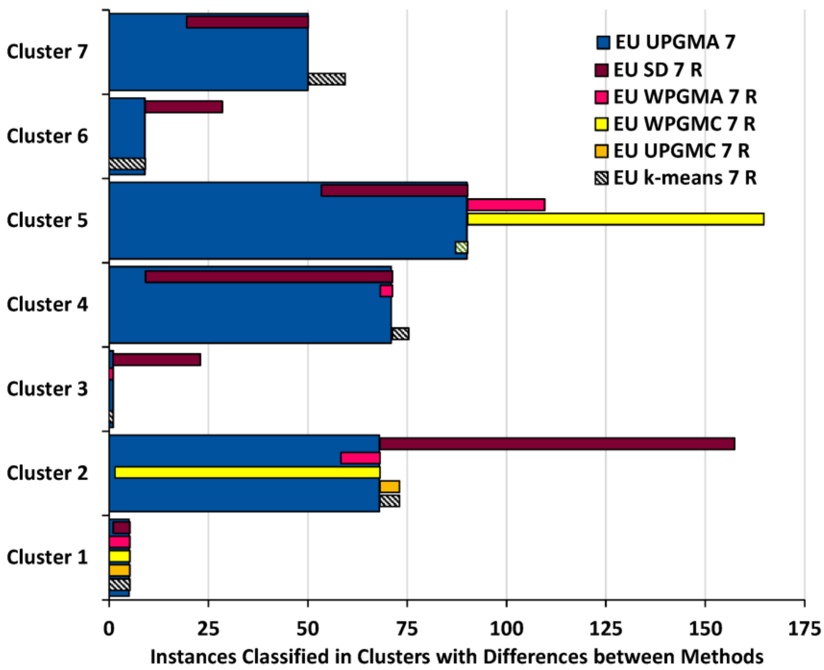

Figure 2: Comparison of the results of different clustering methods for the EU external conditions (EC1, described in Section 3.1) shows the instances classified to the clusters (large bars for UPGMA) and differences between the methods (small bars).

If, following completion of these post-processing analyses, the overview of the situation is clearly understood then the development of a retrofit strategy can commence. Fundamentally, a retrofit strategy consists of the definition of a set of measures applied to a group of buildings connected with the information in which period the application is planned. The definition of this information allows the reduction of expected emissions to be determined for a given investment in the future development of the building stock, which offers very valuable information for policy and decision making.

3. The case study Zernez Energia 2020

3.1 Aim of building stock retrofit and prerequisites

The proposed approach is demonstrated by means of the case of Zernez Energia 2020, an applied research project for the transformation of a Swiss alpine village towards zero emissions in building operation. The aim of the research project is to develop an action plan for a transformation strategy so the building stock and energy systems become emission free in the near future. A combination of retrofitting the existing building stock, transformation of energy systems for heat, electricity, and a model for urban development strategies and sustainable growth of the community including an integrative planning process should enable the municipality to develop and apply measures and policies. Therefore, the project comprises three research modules, each with a specific focus on buildings, energy systems and planning respectively.

The scope of the first module is to develop and establish environmental and economical effective retrofit strategies for the building stock. Since the financial means of building owners and the support by the municipality are limited it is crucial to identify which measures are most effective for building retrofit.

The setting of the project creates a possibility for formulating a bottom-up approach based on real building data. Together with the application patterns, the municipality surveyed and compiled a list of more than 50 parameters for each building including the last retrofit, the condition of the building substance, floor area, installed heating and domestic hot water system, consumption data if available, images and plans for easy recognition and context. A building database was established to quickly access all building parameters for further analysis and calculation. A user interface for the database was created as a centralized building information system for further application by the municipality e.g. energy consulting. Each building then was linked to an identification number, 3D-modelled and inserted into a GIS model for analysing photovoltaic potential (e.g. on roofs). This data serves as basis for the building-detailed analysis and performance-based clustering, which is discussed in detail in Section 4.

External conditions

For the analysis of the building stock and the development of appropriate strategies, external conditions are crucial. The most important external condition (EC) is the CO_2 emissions for electricity from the grid. The determination of the correct value for emissions per kWh is difficult because of the selection of the system boundary. Two plausible scenarios exist: Firstly, in Zernez the majority of the building stock is connected to the European electricity grid ENTSO-E and the consumed electricity in the buildings results in the average emissions of this network, which are 459 g CO_2 per kWh_{el}. This forms the set of external conditions EC1. Secondly, despite the grid network connection, it is possible to consider virtual consumption of energy from a specific source. The source for Zernez is a hydropower plant nearby the village. A specific contract of the village with the Engadiner Kraftwerke AG assigns electricity from the hydropower plant Ova Spin to the village with emissions 4.4 g CO_2 per kWh. This forms the set of external conditions EC2. However, by assigning this low-emission energy to the village withholds it from other consumers, who would need to use other electricity sources with higher emissions. Nonetheless, one can argue that the demand for this low-emission energy would enforce investments in such technology in future. Therefore, both perspectives have their legitimacy and will be considered in this paper. This allows the influence of external conditions to be examined on the clustering results. Whilst the individual condition of the building stock determines the clusters to a significant factor, external factors are one of the most important variables, which is proven later. The conversion factors for the two scenarios are compiled in Table 1.

Table 1: CO_2 conversion factors first considering electricity consumption by EU electricity network ENTSO-E and for the specific contract situation in Zernez.

Type of Heating System	External conditions EC1: EU network kg CO_2 eq. per kWh	External conditions EC2: Zernez kg CO_2 eq. per kWh
Conversion factor oil	0,290	0,290
Conversion factor electricity	**0,459**	**0,00443**
Conversion factor DH	0,045	0,045
Conversion factor wood	0,017	0,017
Conversion factor wood chips	0,022	0,022
Conversion factor heat pump (air/water)	0,061	0,061
Conversion factor heat pump (water/water)	0,066	0,066
Conversion factor heat pump (soil/water)	0,066	0,066
Conversion factor District Heating	0,045	0,045

3.2 Analysis of the current situation

A first analysis of the buildings in this database revealed that there are significant variations in the energy consumption according to the year of construction. It is not the oldest buildings that have the highest energy consumption but the ones constructed in the 20th century, as shown in Figure 3 left top. Furthermore, the exemplary examination of the reaction on a combination of measures (Figure 3, right) illustrates the heterogeneity resulting from the age classification. Consequently, a more effective approach than a standard type-age classification for grouping and strategy development was necessary.

4. The application of performance based clustering

The aim of the clustering is to develop groups of buildings that respond in a similar manner to the same retrofit measures in terms of the reduction of emissions per investment. The basis for the clustering is the response of the individual buildings to selected measures and measure sets. The clustering identifies similarities in the responses and groups the buildings to clusters with a specific characteristic. This characteristic is directly based on the impact that the measures have instead of indirect indicators such as building type, age or other parameters. The procedure follows the methods described in Section 2. Data preparation was outlined in Section 3. The data was then pre-processed in order to determine the effects of the selected measure sets and avoid, for example, runaways leading to one-building clusters (Section 4.1). Subsequently, clustering serves to derive groups (Section 4.2). After the clustering, post-processing consists of statistical analysis, interpretation of the results and development of a transformation strategy (Sections 4.3).

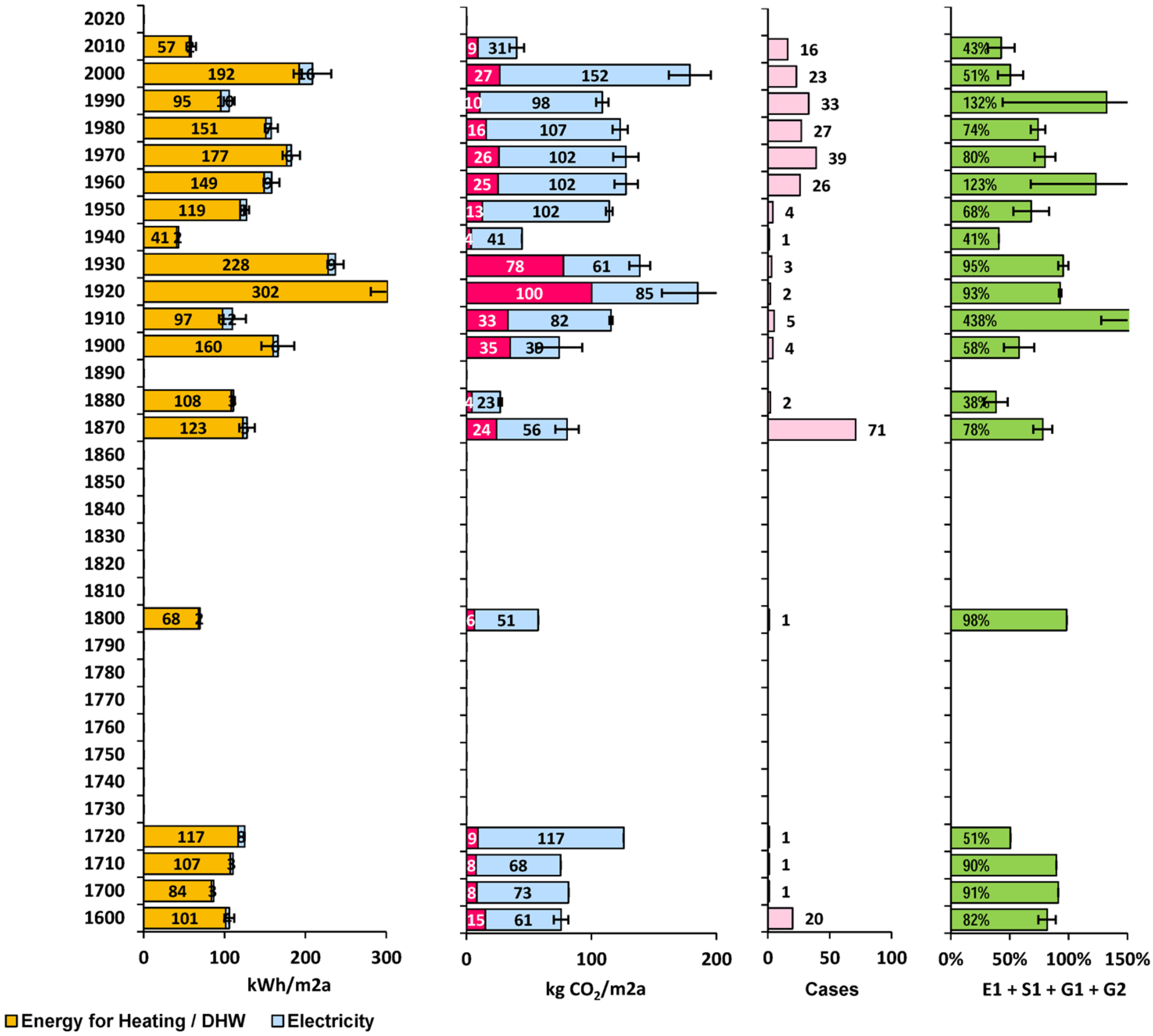

Figure 3: Data analysis of building stock of Zernez by construction year following external conditions EC2 (Zernez conversion factors). Column 1: Energy consumption for heating and electricity; Column 2: CO_2 emissions connected to energy consumption; Column 3: Number of cases / buildings; Column 4: Possible emission reduction by a reference scenario explained in Section 4.

4.1 Pre-processing: measures, impact and scaling

Six different retrofit measures for reducing CO_2 emissions in operation were defined. These were selected according to best practice examples of successful retrofits of the buildings in the village and include measures to reduce energy demand as well as measures to harness locally produced renewable energy. The selected measures include:

- E1: Insulation of the envelope to economic standard;
- E2: Insulation to low-energy standard;
- S1: Replacement of the heating system by a heat pump;
- S2: Connection to the district heating (DH) network;
- G1: De-centralised electricity production by a photovoltaic system;
- G2: De-centralised heat production by a solar thermal system.

S21 is the combination of S1 and S2, which means that the heating system is replaced by a connection to DH if available; otherwise a heat pump is used. An air-water system was chosen as heat pump technology because the investment cost comparison indicated that this is the most cost-effective technology. Furthermore, the combination G12 means the use of a hybrid photovoltaic thermal (PVT) collector. Figure 4 provides an overview of the measures, their combinations and constraining rules.

Measures for reducing CO_2 emissions

The basis for the clustering is an automated analysis of the effects of the measures and the selected combinations on each building, its energy demand and its CO_2 emissions. This analysis uses the detailed database of the current constitution of the buildings and their energy consumption described in Section 3.1. Each measure was executed for each building in the database using the descriptive parameters of geometry, state of construction and heating system and further constraints, such as historic protection etc. Simplified energy calculations described in the following section serve to determine the reduction of the total CO_2 emissions per building by each measure, which forms the effect matrix.

Calculation of emission reduction by measures

The simplified relative energy calculations that use the monitored energy consumption in the database allow the emission reduction to be determined. It is important to note that clustering only requires comparative indicators and is robust in terms of underlying data. Consequently it is not necessary to determine the absolute energy demand in future but the relative effect the measures have.

The simplified calculations are based on geometric information from the database, such as gross floor area, roof form and area, orientation etc., information on the construction, such wall construction, window area and types etc., and on the energy consumption for heating and the electricity consumption. First, the split between energy for heating, for domestic hot water and for electricity consumption is determined. For any cases in which data is missing, assumptions on typical distribution were used. The determination of the split in consumption usage is required to assess the impact of the measures that, for instance, only improve the building envelop on the total building emission. As a next step, the estimation of the missing parameters of the building geometry takes place resulting in the following set of parameters: gross floor area, the areas of the envelope, areas of the façade, windows, wall, roof and foundation as well as roof area available for photovoltaic and solar thermal systems.

For the calculation of the reduction of energy consumption and emissions, the final energy according to DIN 18955-5 is taken into account. This is the energy transferred at the connection between building and network or the delivered fuel for a building's heating system. As the current consumption available in the database includes the proportion of room heating, domestic hot water and electricity, relative changes by the measures are calculated. The results are factors that describe the reduction of energy consumption and emissions.

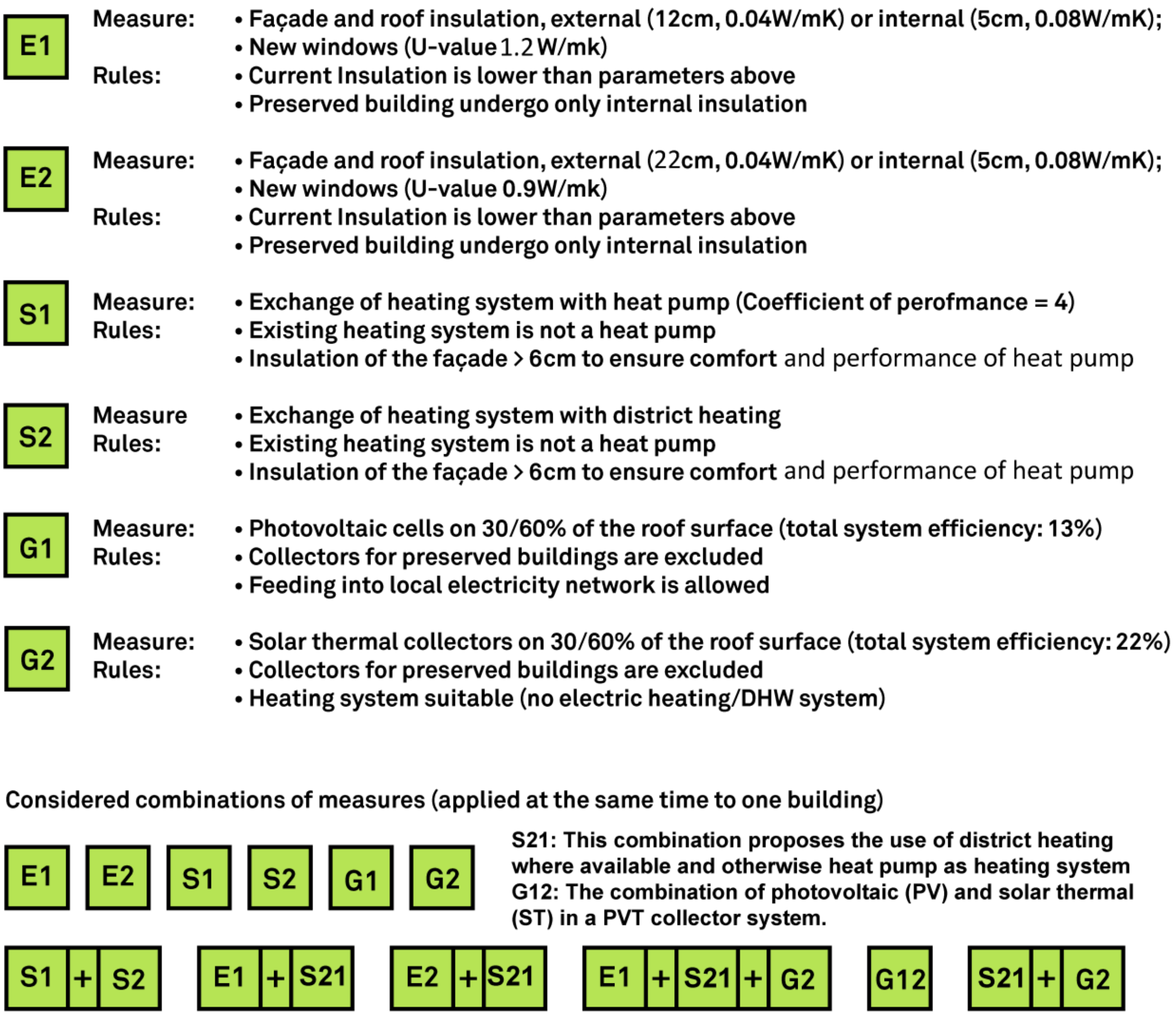

Figure 4: Overview of the measures with constraining rules and their combinations for emission reduction.

For scenarios, which include the replacement of the heating system (measure S1 and S2), the factor of the change of energy demand and emissions to be expected $f_{fhw,sys}$ is derived from old final energy demand $Q_{fhw,old}$ and the new one $Q_{fhw,new}$, which only relies on the ratio of the old system efficiency η_{old} and the new one η_{new}:

$$f_{fhw,sys} = \frac{Q_{fhw,new}}{Q_{fhw,old}} = \frac{\eta_{old}}{\eta_{new}} \text{ with } \eta = \frac{Q_h}{Q_f}. \tag{11}$$

The efficiency is defined as the ratio between usable heat Q_b and the required final energy and includes the coefficient of performance (COP) of heat pumps as well as the efficiency of combustion systems and heat distribution systems.

The factor for change of emissions $f_{CO_2,sys}$, used when switching the energy carrier results from the conversion factors $k_{CO_2,new}$ and $k_{CO_2,old}$ with the conversion factors shown in Table 1:

$$f_{CO_2,sys} = \frac{k_{CO_2,new}}{k_{CO_2,old}} f_{fhw,sys}. \tag{12}$$

Furthermore, the application of the PV system (measure G1) generating the energy amount $Q_{fe,pe}$ leads to the reduction factor of the electric energy demand $f_{fe,pv}$ in relation to the electric energy consumption $Q_{fe,pv}$:

$$f_{fe,pv} = 1 - \frac{Q_{fe,pv}}{Q_{fe,el}}. \tag{13}$$

This calculation assumes an exchange of electricity through the grid so that all electric energy is consumed. Therefore, negative values of $f_{fe,pv}$ are permitted. As the electricity from a PV system has lower emissions than electricity from the grid, which it has replaced, the benefit for emissions $f_{CO_2,el}$ results from the conversion factor for grid electricity $k_{CO_2,grid}$ and for PV $k_{CO_2,PV}$:

$$f_{CO_2,el} = 1 - \left(1 - \frac{k_{CO_2,PV}}{k_{CO_2,grid}}\right)\left(1 - f_{fe,PV}\right). \tag{14}$$

Similarly, the benefit for a solar thermal (ST) system (measure G2) can be calculated. However, in contrast to PV, no replacement by a network is allowed between the buildings and only up to 60% of the final energy can be replaced due to the time shift between availability and demand of heat. Therefore, the reduction by ST is limited to between 0% and 60%. With these limitations, the factor for solar thermal systems $f_{fhw,st}$ relies on the ratio of the solar thermal energy gain $Q_{fhw,ST}$ and the energy demand Q_{fhw}:

$$f_{fhw,ST} = 1 - \max\left(\min\left(\frac{Q_{fhw,ST}}{Q_{fhw}}, 0.6\right), 0\right) \tag{15}$$

For the energetic retrofit of the component i of the building envelope (measure E1 and E2) the factor describing the change of energy demand $f_{fhw,comp,i}$ results from the share that component has in the heat flow:

$$\begin{gathered} f_{fhw,comp,i} = 1 - a_{ht} f_{AU,i}\left(1 - f_{U,i}\right) \text{ with } \sum_{i=1}^{n_{comp}} f_{AU,i} = 1, \\ a_{ht} = \frac{Q_{fh,old}}{Q_{fhw,old}},\ f_{AU,i} = \frac{U_{comp,i,old} A_{comp,i}}{U_{tot,old} A_{tot}} \text{ and } f_{U,i} = \frac{U_{comp,new}}{U_{comp,old}}. \end{gathered} \tag{16}$$

The factor a_{ht} describes the transmission share of the heat demand for rooms and water in terms of final energy. The factor $f_{AU,i}$ describes the share of area in the buildings envelope weighted by the U value of the component as part of the total area of the envelope A_{tot} and the average U value $U_{tot,old}$. The factor $f_{U,i}$ describes the ratio of the improvement of the U value of the component i. The combination of the improvement of n components is

$$f_{fhw,comp,tot} = \sum_{i=1}^{n_{comp}} f_{fhw,comp,i}. \tag{17}$$

Furthermore, embedded energy in the components of the envelope and in technology components is considered. For heating systems and collectors, the conversion factors include these effects. For the envelope, a separate calculation is required. Therefore, the factor for CO_2 emission needs an adaptation for embedded energy:

$$f_{CO_2,comp} = 1 - \sum_{n=1}^{n_{comp}} \max\left(1 - f_{fhw,comp,i} - \frac{V_{comp,i} k_{CO_2,comp,i}}{Q_{fhw} k_{CO_2}}; 0\right). \tag{18}$$

$V_{comp,i}$ is the volume of the component's material and $k_{CO_2,comp,i}$ the conversion factor for emissions for the component production. The volume is determined from the database, the geometry and the definition of measures above. The emission factor results from the ecoinvent database (www.ecoinvent.ch). Furthermore, it assumed that a measure that results in higher emissions caused by embedded energy than saved emission during operation is not applied.

All changes by measures result in the total reduction factor $f_{fhw,tot}$ for heating and hot water and f_{fe} for electricity:

$$f_{fhw,tot} = 1 - \left(1 - f_{fhw,sys}\right)\left(1 - f_{fhw,ST}\right)\left(1 - f_{fhw,comp}\right) \text{ and } f_{fe} = f_{fe,PV}. \tag{19}$$

Furthermore, the change of the heating system has to be considered for the CO_2 reduction factor $f_{CO_2,fhw,tot}$:

$$f_{CO_2,fhw,tot} = f_{CO_2,sys} f_{fhw,tot}. \tag{20}$$

Moreover, this calculation considers several constraints, e.g. that external insulation is not applicable to a historic façade or that a heat pump system with low temperature radiators does not work in conjunction with a badly-insulated façade; this is the implementation of the constraining rules shown in Figure 4.

Potentials of the measures

Figure 5 shows the results of the measures' and sets' potentials for both sets of external conditions, electricity from the EU grid (EC1) and the local hydropower plant (EC2). The automated analysis determines the applicability of all measures for all buildings in the database and their impact. A measure is applied to a building in the database if the criteria and constraints mentioned above are met. The most effective single measure according to this analysis is replacement of the building's heating supply by a heat pump (S1) with a reduction of 27.1% of the CO_2 emissions (EC1) and 53.4% (EC2). Applying a set of measures, such as insulation plus system replacement plus solar thermal collectors for EC1 (E1 + S21 + G2) or insulation plus heating system replacement (E2 + S21) for EC2, leads to the highest observed reductions of 50.1% (EC1) and 84.5% (EC2); however, this requires high economic investments. Other measures or sets have a much better efficiency in terms of avoided emissions per investment cost, such as S21, which is the replacement of the heating system by district heating, if available, otherwise installation of a heat pump. Furthermore, there are cases in that investing more and applying more measures leads to poorer results than more cost effective options; e.g., in the EU external conditions (EC1), the application of low-energy insulation with heating system replacement (E2 + S21) is more expensive but less effective than the use of reduced insulation with heating system replacement and solar thermal collectors (E1 + S21 + G2). This illustrates the vital need for specifically tailored strategies. The set ‘Clustered’ shows the reduction based on measures of such specifically developed strategies for the clusters described in the next section. In almost all cases, energy consumption decreases. However electricity consumption for EC2 increases as oil systems are replaced by heat pumps. In general, the clustering strategies are a Pareto trade-off with the objective to identify the lowest-cost investments and to reduce the emissions as much as possible.

4.2 Clustering

In the next step, hierarchical distance-based clustering serves to identify similar responding cases in the compressed effect matrix E_{comp}. The basis is the Euclidian distance of the cases in the feature space that forms a similarity matrix. In this phase, two different methods of hierarchical clustering were used: Unweighted Pair Group Method with Arithmetic Mean (UPGMA) fuses the clusters whose coordinate-average centroids are closest and Shortest Distance (SD) fuses the clusters with the closest individuals. Both methods lead to well-separated clusters in case of the second set of external conditions EC2, i.e. the local electricity scenario. However, in case of the EU external conditions EC1 only the SD method, which is intrinsically trimmed to maximize separation, delivered usable results in terms of cluster separation. The clustering and the partial post-processing, shown in this and the next section, was implemented in MATLAB using the Statistics Toolbox.

Dendrograms helped to determine the number of clusters by experimenting with different numbers of clusters and interpretation on a trial basis. For instance, the dendrogram shown in Figure 6 was used to select the number of clusters (7) as the distance between ramifications increases after seven clusters. Further analysis and interpretation of the seven found clusters, which is described in Section 4.3, lead to clear characterisation, which supports this selection. Furthermore, this was executed until a consistent interpretation of the clusters was possible and the number of clusters was small enough to be managed. This testing also included the manual merging of clusters with only a few members into larger clusters.

Scatterplots were utilised to identify important measures that lead to cluster separation and thus are important strategic elements. Figure 7 shows all pairwise combinations of scatterplots for EC1, which is EU electricity. In these plots, Box 1 served to identify the difference between the clusters 1.1 to 1.4 that is reproduced larger in Figure 8 (top) and interpreted in the text. Furthermore, Box 2 served to identify the three core groups of reaction to solar collectors (PV and ST), which is shown in detail in Figure 8 (bottom).

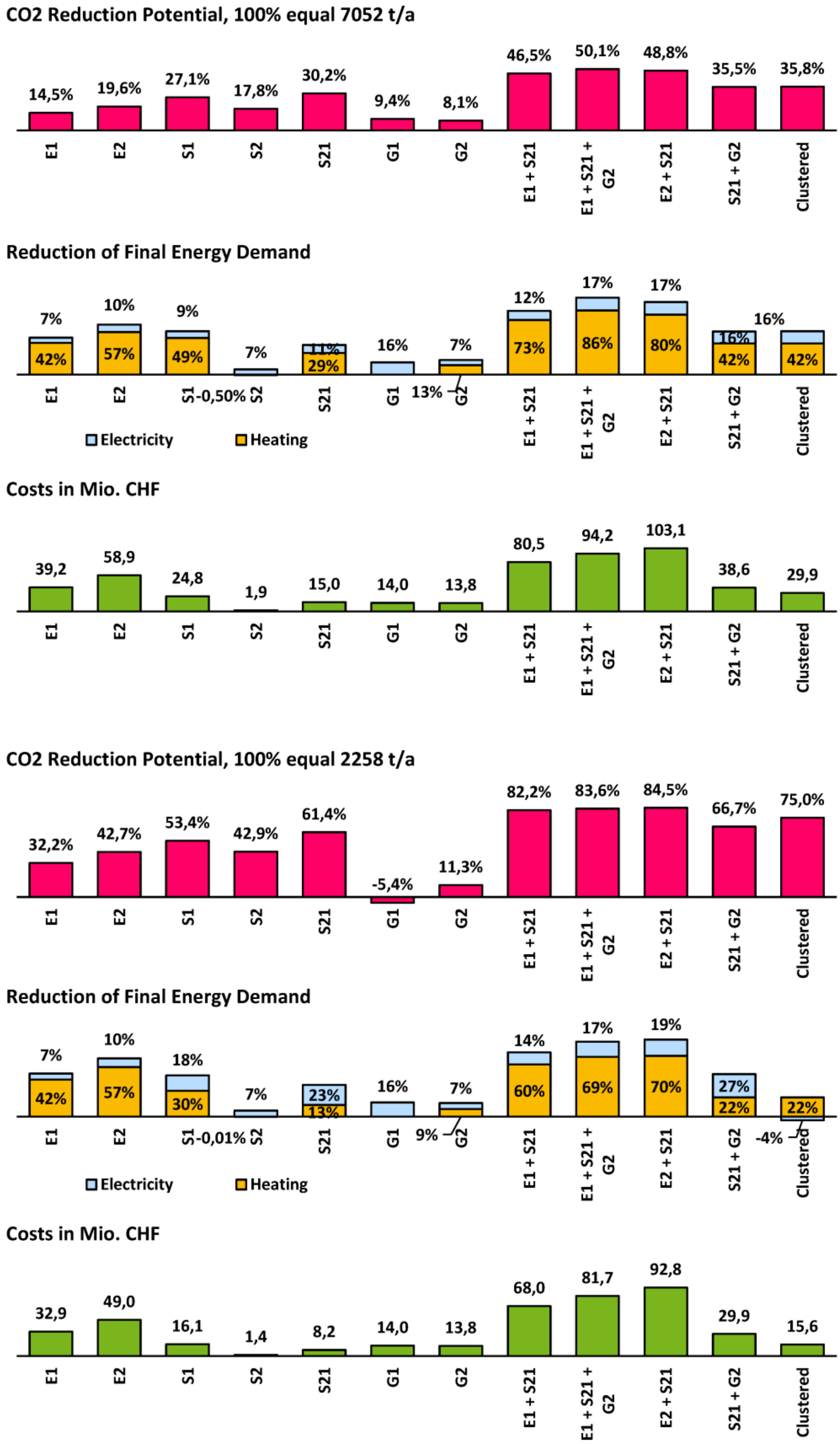

Figure 5: Effect of the measures and measure sets, top three charts: for external conditions EC1 (EU electricity), bottom three charts: external conditions EC2 (local electricity)

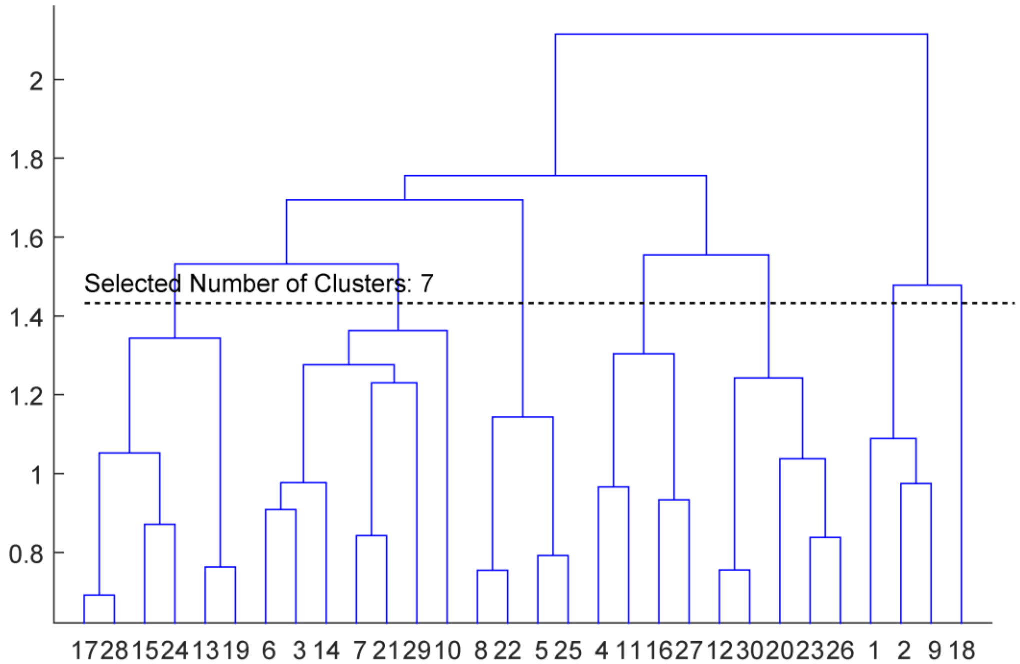

Figure 6. Dendrogram for the EU data (EC1).

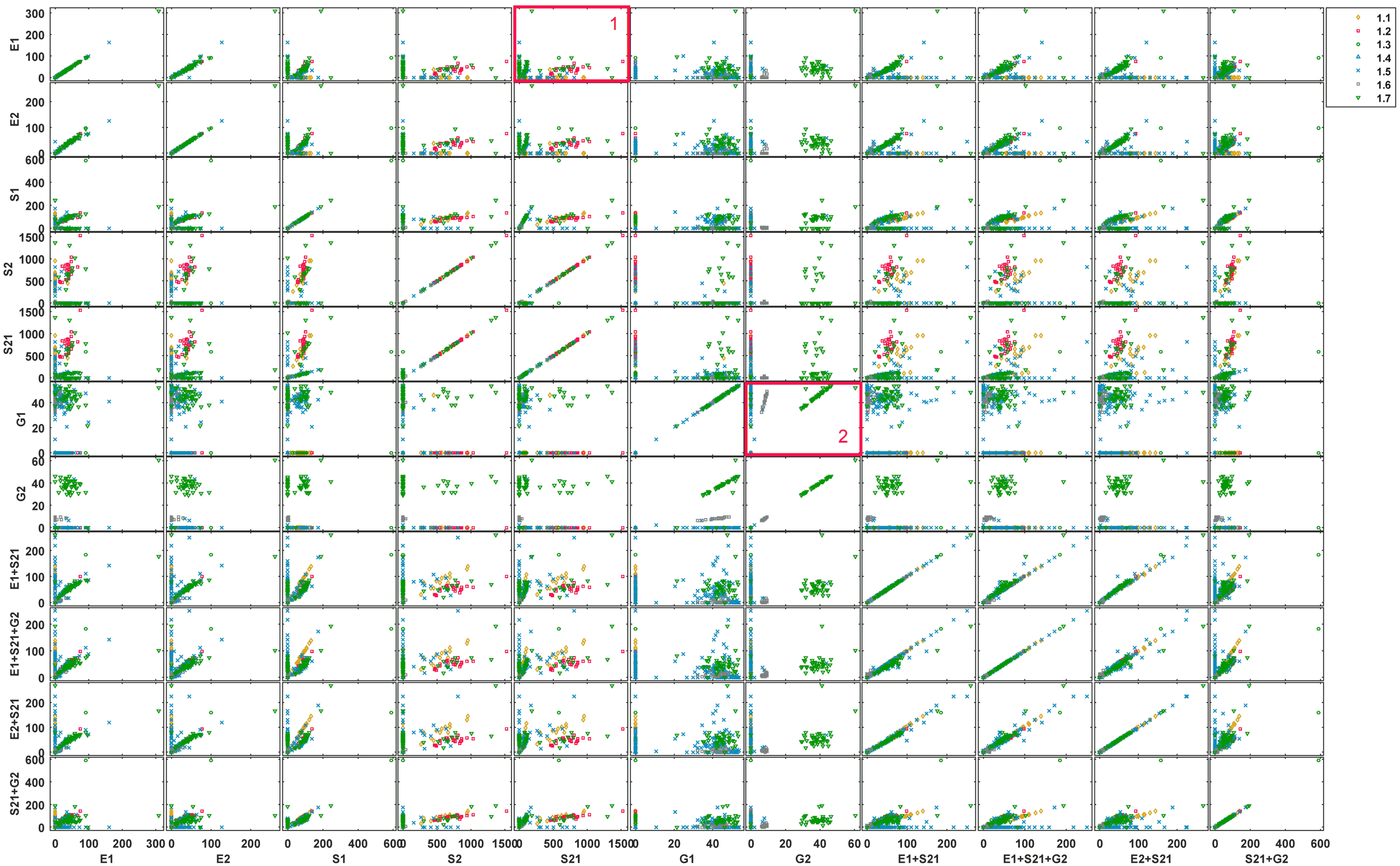

Figure 7: Scatterplots of the four clusters in the nine-dimensional feature space for the data based on external condition EC1 (EU).

4.3 Post-processing and results

The first step of post-processing the clustering results is the interpretation of the clusters. This starts with the visual analysis of the scatter plots. Figures 8 and 9 show selected scatterplots of the EU and the local grid results comparing the response of the building stock to different retrofit measures; each dot represents one building and the axis compare two selected measures or measure sets. The plots locate buildings with similar reaction to the measures close to each other, forming a point cloud, thus allowing the identification of clusters. However, closeness is defined, as describe previously, in the nine-dimensional feature space; selected viewing in two dimensions represents a way of accessing the cluster information. The location of the point clouds gives indications on the characteristic of the clusters. In the following, this is illustrated for the five identified clusters, the measures and their combinations:

Cluster Characteristics for external conditions EC1 (EU electricity mix, Figure 8):

Cluster 1.1: The CO2 emissions of the buildings in this cluster show only little reaction to the improvement of insulation (E1); but responds to the replacement of the heating system (S21). Further parameter analysis has shown that the cluster includes older buildings with large floor area that mainly use electric heating. Therefore measure S21 is most cost-effective, which is the reason for selecting this measure.

Cluster 1.2: In this cluster, oil heating systems with high CO_2 emissions dominate; therefore, the buildings strongly respond to insulation and heating system replacement. The cluster includes large non-detached buildings similar to Cluster 1.1 that are mainly constructed during the 19th century. For this cluster, measures S21 and E1 are most cost-effective and thus selected.

Cluster 1.3: In contrast, buildings in Cluster 1.3 strongly respond to insulation improvement but heating system replacement yields less successful results. This cluster includes mainly small detached houses with oil or electric heating systems. As a result, measure E1 and E2 are most cost-effective.

Cluster 1.4: This cluster responds to none of the retrofit measures. The reason is that the buildings are either already connected to district heating with low emissions or have no heating system installed. Therefore, no measures are applied.

Cluster 1.5: Clusters 1.5 to 1.7 include buildings that respond well to solar thermal collectors (ST) and photovoltaic collectors (PV). Most of these buildings are detached houses that have an inclined roof offering a well-suited area for the placement of ST or PV. The buildings in Cluster 1.5 only respond to PV as they either have no heating or the heating is not compatible with a solar thermal support, such as wood heating. Only in combination with measure S21 (replacement of the heating system) the measure G2 (ST) is applicable, as defined by the rules for the measures. The application of measure G1 (PV) and, given an existing heat demand, measures S21 + G2 are most cost-effective.

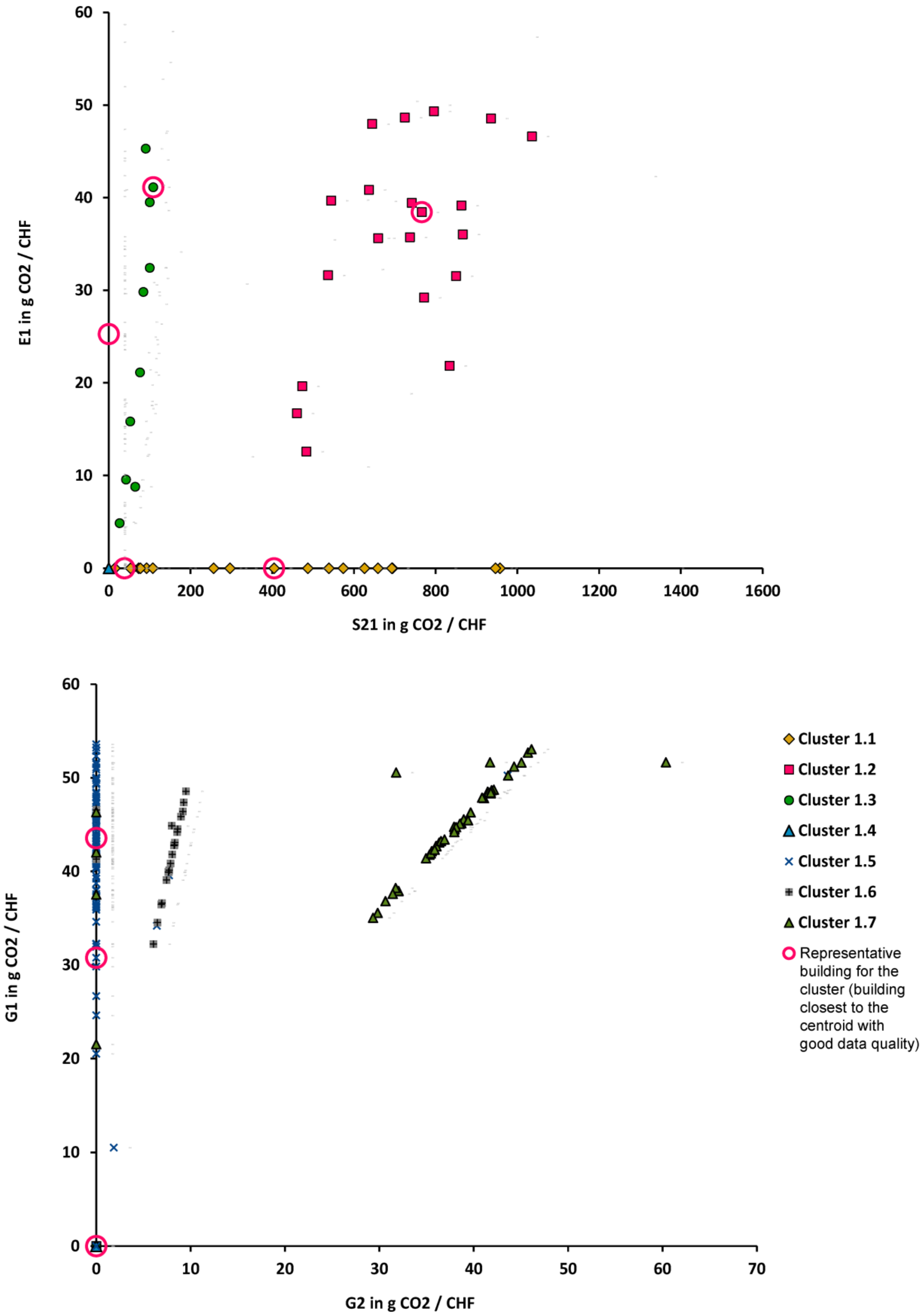

Figure 8: Two scatter plots showing the clustering for the data set based on external conditions EC1 (EU); Top: Clusters 1.1 to 1.4 for the measures S21 and E1; Bottom: Clusters 1.5 to 1.7 for the measures G2 and G1.

Cluster 1.6: This cluster includes buildings that respond well to the placement of PV and slightly to ST systems installation. The reason is that the buildings already use heat pumps and therefore the emissions per kWh heat that would be replaced by ST are already reduced. For this reason, only G1 (PV) is cost-effective.

Cluster 1.7: This cluster comprises buildings that respond well to both types of collectors. The buildings in this cluster are mainly heated by oil; therefore, all substitutions of consumed energy results in reduced CO_2 emissions. Therefore, the change of heating system such as measure S21 + G1 + G2 is optimal. In all cases that comprise both measures, G1 and G2 hybrid collectors (PVT) combining PV and ST are used.

Cluster characteristics for external conditions EC2 (local electricity mix, Figure 9):

Cluster 2.1: This cluster contains buildings with very high CO_2 emissions, mainly caused by large heating energy demands that are supplied by oil heating. These buildings respond very well to insulation measures (E1 / E2) and the replacement of the heating systems (S21). However, S21 is far more cost-effective than E1; therefore, only S21 is selected.

Cluster 2.2: This cluster comprises buildings with a smaller heating energy demand that also employ oil heating. Even though they are less well insulated as the buildings in Cluster 2.1 they respond less well to retrofit measures. The reason is a more compact building shape or more economic user behaviour. However, S21 is very cost-effective and thus selected.

Cluster 2.3: In this cluster mainly small buildings with medium energy demand for heating and emissions are grouped. As in Cluster 2.1 and 2.2 these buildings employ oil heating. Similarly to Cluster 2.1 they respond well to insulation measures and replacement of the heating system; however due to the smaller demand, heating system replacement is a little less cost-effective. Nonetheless, both measures, insulation and heating system replacement show a favourable level of cost-effectiveness and, thus, are selected.

Cluster 2.4: This cluster contains buildings with very little or no CO_2-emissions mainly because they do not require heating or they already use heating systems based on renewable energy sources. Therefore, measures show only little or no effects and are thus not cost-effective.

Cluster 2.5: This cluster contains few special buildings with high CO_2-emissions that are singular within the building stock. Their characteristics vary; due to the high CO_2-emissions, they respond well to retrofit measures. However, insulation (E2) and heating system replacement (S21) are most cost-effective and thus selected.

Analysis of the clustering allows the derivation of partial retrofit strategies for each cluster. Tables 2 and 3 give an overview of these strategies for the clusters for both kinds of external conditions. These recommendations represent a trade-off between reduction of CO_2 emissions and investment costs.

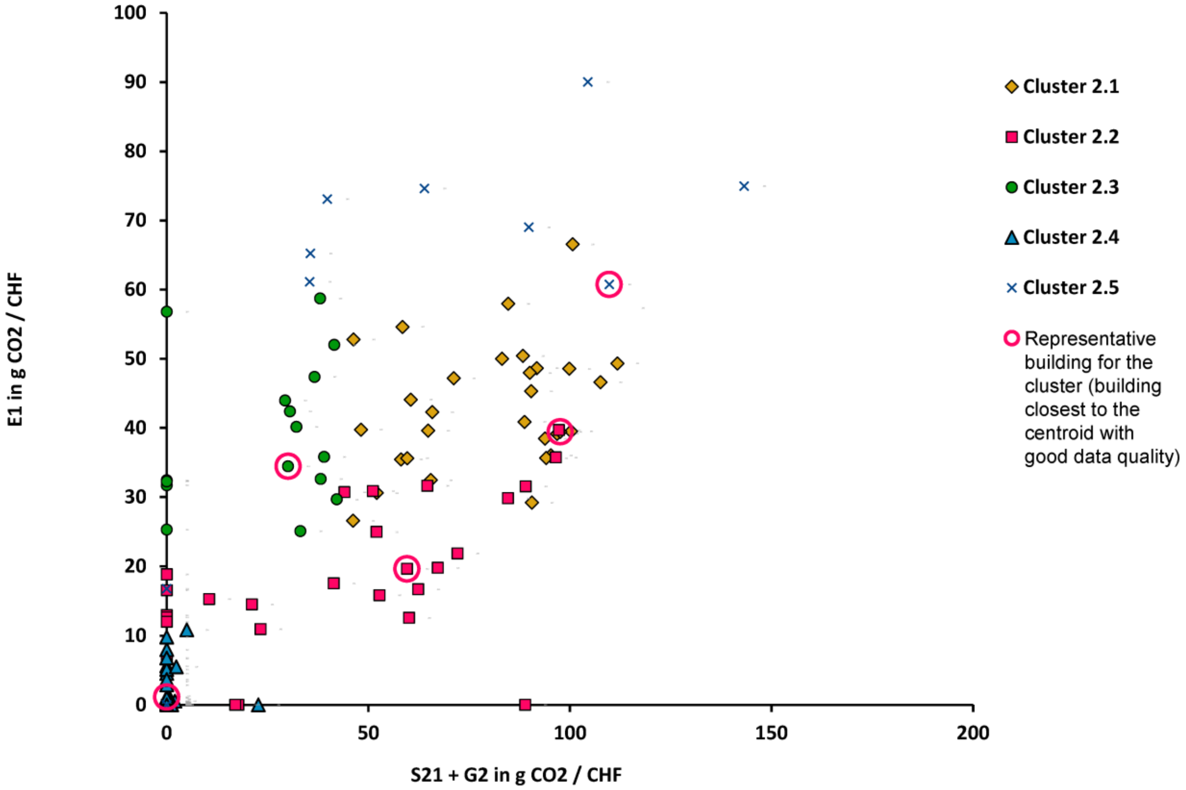

Figure 9: One scatter plot of clustering results for the data set based on external conditions EC2 (local electricity) for the measures S21 + G2 and E1.

Table 2: Partial retrofit strategies derived from the clustering for external conditions EC1.

External conditions EC1: EU electricity

Cluster	Measure Combination
1.1	S21
1.2	S21
1.3	E2 + S21
1.4	None
1.5	S21 + G2 + G1
1.6	G1
1.7	S21 + G2 + G1

Table 3: Partial retrofit strategies derived from the clustering for external conditions EC2.

External conditions EC2: Zernez electricity	
Cluster	**Measure Combination**
2.1	S21
2.2	S21
2.3	E2 + S21
2.4	None
2.5	E2 + S21

To develop and control these strategies, cost-efficiency curves, as shown exemplarily in Figure 10, are an important tool. They originate directly from the building database with its individual cases. For one measure, one measure combination or for one clustering scenario, an algorithm ranks the cases from highest emission reduction per investment descending and sums up emission reduction and costs. The curves show these sums. By following the curves, it is possible to see the performance of a measure combination or scenario in order to ensure that the retrofitting commences with the "lowest hanging fruits". This allows investments in the buildings in which measures have the largest effect compared to their costs. The most desirable curve have a flat ascent and go as far as possible to the right in Figure 10.

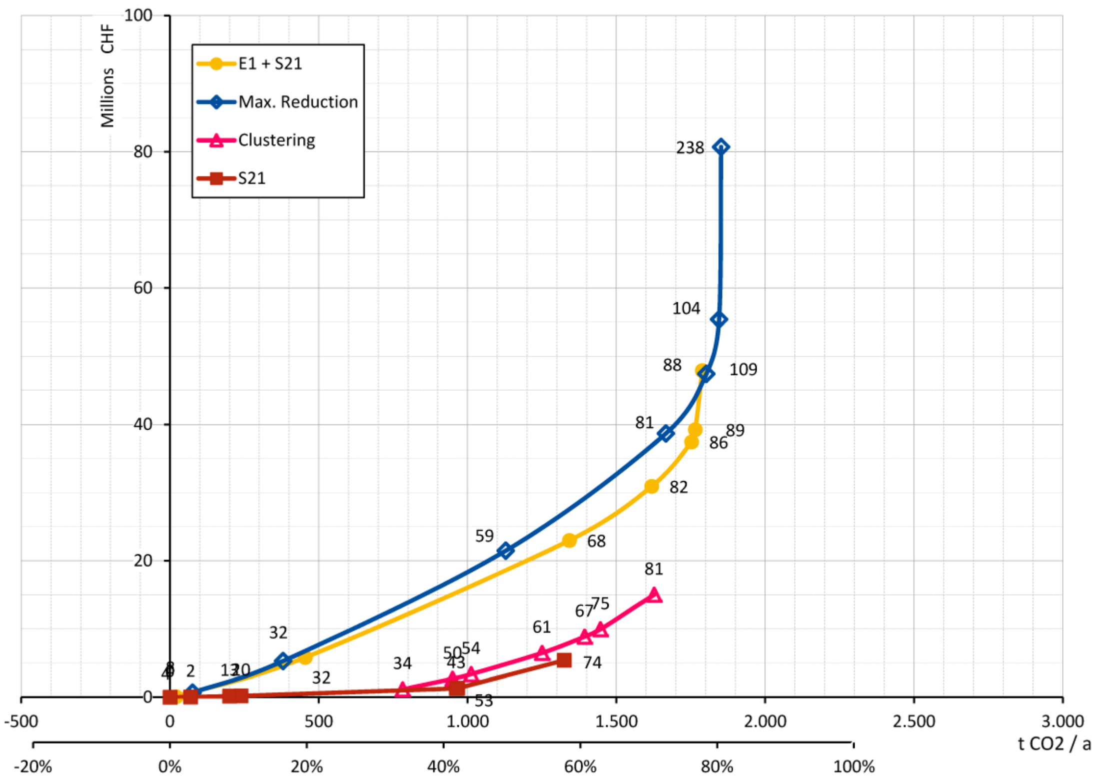

Figure 10: Selected cost-efficiency curves for external conditions EC2 (Zernez electricity); the numbers at the data points indicate the count of retrofitted buildings.

Deriving transformation strategies

To provide the village Zernez with a policy for a low-emissions village, the approach allowed the partial retrofit strategies to be compiled into an overall transformation strategy. This included the consideration of yearly available investments for energy efficiency and emission reduction in the building stock and therefore developed a time-based transformation of the building stock to a low-emission state.

This transformation is based on a sequence of clusters because a strategy referring to a one-by-one sequence of buildings, as provided by the cost-efficiency curve in Section 4.3, is not implementable as a retrofit policy for large building stocks. In correlation to the cost-efficiency curves, the transformation strategy "Clustering" shown in Figure 11 starts with the clusters that have the best ratio of CO_2 emission reduction per investment. In case of the Zernez conditions, this is Cluster 2.1 followed by Clusters 2.2, 2.5 and 2.3. Table 4 shows the order and the average efficiency of the selected strategies. Cluster 2.4 buildings are only touched in the context of regular retrofitting, which includes all buildings. By using this intelligent retrofit procedure and handling the "lowest hanging fruits" first, the clustering strategy demonstrates that an investment of 0.6 Mio. CHF per year can achieve a reduction of nearly 80% of the emissions in the next 35 years with a rapid descent over the first few years. The cost-effectiveness decreases after the four selected clusters; further reductions in CO_2 are possible (up to zero emissions of the building stock), these however result in high costs with increasingly less effectiveness. In contrast, investing approximately the same amount equally in all buildings only achieves a reduction of 20% of the emissions by 2050 ("Strategy Equal Investments, 0.5 Mio. CHF/a"). Even doubling the investments whilst following the same strategy leads to a reduction of only 40% of the emissions. The base-line for E1, S1 and E1+S1 is constructed from the historical retrofits included in the database and the assignment to measures defined in Figure 4. These are the scenarios if investments were made in the same way as in the past and no special attention is paid to energy and emissions.

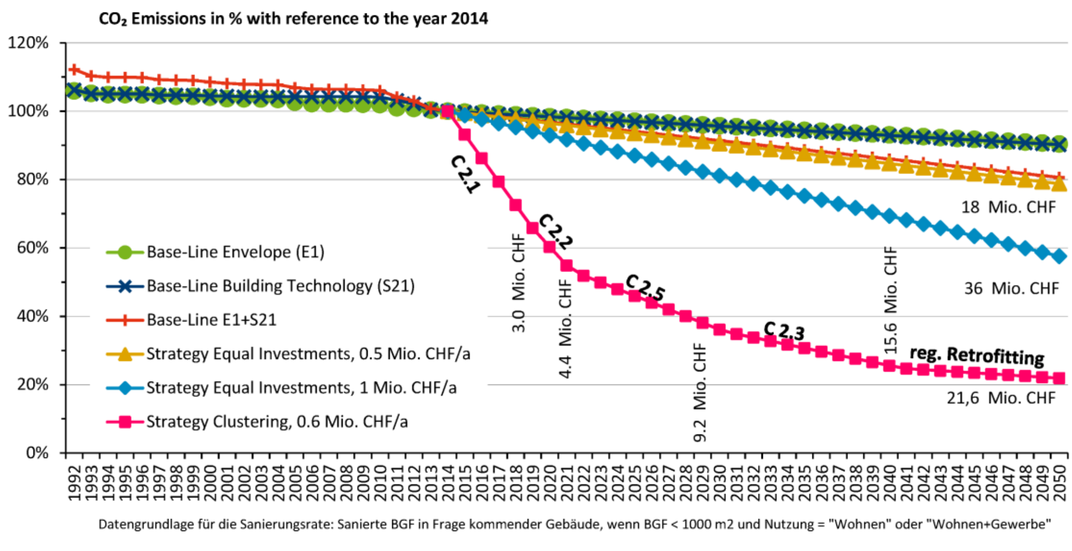

Figure 11: Developed transformation strategy for Zernez with electricity from the nearby hydropower plant (external conditions EC2)

Table 4: Average efficiency of the strategies per cluster for the Zernez conditions (EC2).

Cluster	Strategy	g CO_2 / CHF
2.1	S21	258
2.2	S21	205
2.5	E2 + S21	77
2.3	E2 + S21	40
2.4	None	0

4.4 Interpretation

The exemplary developed transformation strategy for the case study demonstrates the benefit of a clustering approach that is based on measure efficiencies. An 80% reduction in building stock emissions makes a significant contribution to the project target of zero carbon. Further contributions of the energy systems that are not described here allow the project target to be achieved.

Instead of choosing just one descriptive parameter of a building to define a retrofit (such as type or age), the proposed method allows incorporation of a range of descriptive parameters determining the effect of retrofit measures and the explicit effect of such measures. The identification of groups of buildings that react similarly on measures allows a retrofit strategy to be developed for multiple buildings simultaneously and that focuses on those buildings that provide the best benefit per investment without needing to consider each building individually which is often not possible due to the number of buildings involved. This group identification is highly valuable for developing retrofit strategies that can form the basis for investment decisions, for portfolio management or for policy making as in the case for Zernez Energia 2020.

Using surveyed building data instead of statistical data bears a certain risk of false entries due to a lack of knowledge of the person asked or asking. A database with subsequent deviations of stored data versus real data might lead to skewed results. As the method focuses on describing the characteristics of groups of buildings with similar properties rather than relying on the detailed description and result for an individual building, occasional false entries of single building parameters are not as critical. Although the method determines the potential for each measure per individual building, the results for the cluster do not necessarily represent the optimal measures at the level of each individual building. The method aims to identify groups of buildings and their potential with respect to the measure combinations. The retrofit of a single building requires the detailed examination of the building, which is not feasible in the scope of such a strategic alignment for a building stock of more than 300 buildings.

Furthermore, the two different definitions of external condition sets EC1 and EC2 illustrate that such conditions have a high influence. Consequently it is essential to properly choose and rationalise any assumptions, for both the application of clustering and for any types of assessment of CO_2 emissions of buildings or buildings stocks. In the development of the project, it has generally been observed that clustering is robust against small bias in the data. Slightly changed datasets had minimal effect on the clustering. The general strategies were also valid for such modified datasets.

Finally, the data analysis of the building stock, with respect to the classification by the clusters in Figure 12, shows that the clusters provide a far better classification for the purpose of retrofitting than the building age shown in Figure 3. The clustering leads to nearly no deviations in the reaction on the measure E2 + S21 (Figure 12, diagram on the right), which is one of the important retrofitting measures whereas the reactions for the age classification show high deviations (Figure 3, right). Additionally, clustering results in only 5 groups whereas usual age grouping (e.g. IWU 2012) use 15 to 50 categories. This makes strategy development and management in policy making less complicated and more effective.

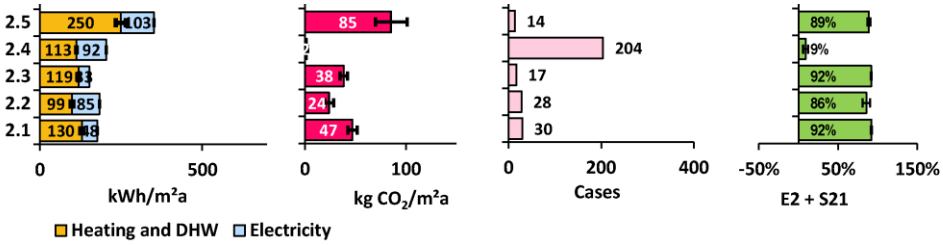

Figure 12. Characteristics of the clusters in terms of energy demand, emissions and number of buildings and their response and thus effectiveness to the most effective measure combination E2+S21.

5. Conclusions

The objective of this work was to use readily available rich building data for the simulation, analysis and identification of cost-effective retrofit measures. Clustering the result data based on the buildings' response to the different retrofit measures delivered groups of buildings with similar cost-effectiveness of specific measures. Post-processing showed that the clustering method led to a very good similarity of energy and emission characteristics as well as reaction on measures.

The automated assessment and the identification of similarities by means of clustering has potential to support the development of retrofit strategies of entire building stocks. The identification of building clusters of similar response and the quantification of their impact in terms of CO_2 mitigation supports the development and classification of funding schemes and policies to guide retrofit measures. For instance, in the case of Zernez, it aided the development of a strategy to transform the building stock to a low-emission status in the near future.

In summary, this approach represents a feasible method for the strategic management for retrofitting a building stock. Focusing on cost-effectiveness of measures not only identifies potential for emission reduction but also considers their economic feasibility and optimizes the application of investments to reduce CO_2 emissions. Even though the method was developed for, and is applied to, a specific case study it can be used on any building stock. Most of its steps can be automated. Only boundary conditions such as the number of clusters and the result interpretation need manual intervention. It is the availability and quality of the building data that sets the constraints for its application. In Switzerland, for example, local authorities, as well as public institutions, are increasingly establishing rich building databases that can be utilized for a wide variety of further data analysis in the future. The case of Zernez thereby provides an outlook of what is possible if such data is available and carefully utilised. Its boundary conditions, mix of building types and CO_2 emissions reduction targets make it a small-scale yet prototypical case study that can be compared to larger building stocks and related strategies.

6. Acknowledgements

This work is partly funded by the Swiss Commission for Technology and Innovation (CTI). We furthermore acknowledge the contributions of the research groups of ETH Chairs of Architecture and Urban Design, Prof. Kees Christiaanse, Michael Wagner (project lead), Building Physics, Prof. Dr. Jan Carmeliet, Dr. Kristina Orehounig and the ETH Institute for Environmental Engineering, Prof. Dr. Stefanie Hellweg, and Niko Heeren, Andreas Frömelt and Bernhard Steubing for their contributions to the building database and data analysis.

References

Boardman B (2007): Examining the carbon agenda via the 40% House scenario, Building Research & Information 35(4), pp. 363-378, http://dx.doi.org/10.1080/09613210701238276.

Forgy EW (1965): Cluster analysis of multivariate data: efficiency versus interpretability of classifications, Biometrics 21, pp. 768-769.

Gaitani N, Lehmann C, Santamouris M, Mihalakakou G (2010): Using principal component and cluster analysis in the heating evaluation of the school building sector, Applied Energy 87(6), pp. 2079-2086, http://dx.doi.org/10.1016/j.apenergy.2009.12.007.

Hernandez P, Burke K, Lewis JO (2008): Development of energy performance benchmarks and building energy ratings for non-domestic buildings: An example for Irish primary schools, Energy and Buildings 40(3), pp. 249-254, http://dx.doi.org/10.1016/j.enbuild.2007.02.020.

Hung WH, Kang SJC (2014): Automatic clustering method for real-time construction simulation, Advanced Engineering Informatics 28(2), pp. 138-152, http://dx.doi.org/10.1016/j.aei.2014.02.001.

Hyun KH, Lee JH, Kim M, Cho S (2015): Style synthesis and analysis of car designs for style quantification based on product appearance similarities, Advanced Engineering Informatics, in press, http://dx.doi.org/10.1016/j.aei.2015.04.001.

IWU (Institut Wohnen und Umwelt) (2012): Typology Approach for Building Stock Energy Assessment. Main Results of the TABULA project, www.building-typology.eu, accessed Dec 2014.

Jazizadeh F, Becerik-Gerber B, Berges M, Soibelman L (2014): An unsupervised hierarchical clustering based heuristic algorithm for facilitated training of electricity consumption disaggregation systems, Advanced Engineering Informatics 28(4), pp. 311-326, http://dx.doi.org/10.1016/j.aei.2014.09.004.

Jones PJ, Lannon S, Williams J (2001). Modelling building energy use at urban scale. Proceedings of the 7th International IBPSA conference, Rio de Janeiro, Brazil, pp. 175—180.

Kavgic M, Mavrogianni A, Mumovic D, Summerfield A (2010): A review of bottom-up building stock models for energy consumption in the residential sector, Building and Environment 45(7), pp. 1683-1697, http://dx.doi.org/10.1016/j.buildenv.2010.01.021.

Kohler N, Hassler U, Paschen H (1999): Stoffströme und Kosten in den Bereichen Bauen und Wohnen - Konzept Nachhaltigkeit, Springer Berlin Heidelberg, http://dx.doi.org/10.1007/978-3-642-58503-6_1.

Kohler N, Yang W (2007): Long-term management of building stocks, Building Research & Information 35(4), pp. 351-362, http://dx.doi.org/10.1080/09613210701308962.

Lloyd S (1982): Least squares quantization in PCM, 28(2), pp. 129-137, http://dx.doi.org/10.1109/TIT.1982.1056489.

Motamedi A, Soltani MM, Hammad A (2013): Localization of RFID-equipped assets during the operation phase of facilities, Advanced Engineering Informatics 27(4), pp. 566-579, http://dx.doi.org/10.1016/j.aei.2013.07.001.

de Oliveira DP, Garrett Jr. JH, Soibelman L (2011): A density-based spatial clustering approach for defining local indicators of drinking water distribution pipe breakage, Advanced Engineering Informatics 25(2), pp. 380-389, http://dx.doi.org/10.1016/j.aei.2010.09.001.

Santamouris M, Mihalakakou G, Patargias P, Gaitani N (2007). Using intelligent clustering techniques to classify the energy performance of school buildings, Energy and Buildings 39(1), pp. 45-51, http://dx.doi.org/10.1016/j.enbuild.2006.04.018.

Simovici DA, Djeraba C (2008): Clustering, in: Simovici DA, Djeraba C: Mathematical Tools for Data Mining, Springer London, pp. 495-525, http://dx.doi.org/10.1007/978-1-84800-201-2_13.

Sokal RR, Michener CD (1958): A Statistical Method for Evaluating Systematic Relationships, Univ. Kansas Sci. Bull. 38, pp. 1409-1438.

Swan LG, Ugursal VI (2009): Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, Renewable and Sustainable Energy Reviews 13(8), pp. 1819-1835, http://dx.doi.org/10.1016/j.rser.2008.09.033.

Uihlein A, Eder P (2010): Policy options towards an energy efficient residential building stock in the EU-27, Energy and Buildings 42(6), pp. 791-798, http://dx.doi.org/10.1016/j.enbuild.2009.11.016.

Xu R, Wunsch DC (2009): Clustering - IEEE series on computational intelligence, Wiley, Oxford.

Yamaguchi Y, Shimoda Y, Mizuno M (2007): Proposal of a modeling approach considering urban form for evaluation of city level energy management, Energy and Buildings 39(5), pp. 580-592, http://dx.doi.org/10.1016/j.enbuild.2006.09.011.